Immersive Teleoperation of the Eye Gaze of Social Robots Assessing Gaze-Contingent Control of Vergence, Yaw and Pitch of Robotic Eyes

Remi Cambuzat, Frédéric Elisei, Gérard Bailly, Olivier Simonin, Anne Spalanzani

Immersive Teleoperation of the Eye Gaze of Social Robots

Assessing Gaze-Contingent Control of Vergence, Yaw and Pitch of Robotic Eyes

Rémi Cambuzat[a,b], Frédéric Elisei[a], Gérard Bailly[a], Olivier Simonin[b,c], and Anne Spalanzani[c]
[a]GIPSA-lab, Univ. Grenoble-Alpes, CNRS, Grenoble INP, Grenoble - France
[b]CITI-lab, INSA Lyon, INRIA, Lyon - France
[c]Univ. Grenoble-Alpes, INRIA, Grenoble - France

Abstract

This paper presents a new teleoperation system – called stereo gaze-contingent steering (SGCS) – able to seamlessly control the vergence, yaw and pitch of the eyes of a humanoid robot – here an iCub robot – from the actual gaze direction of a remote pilot. The video stream captured by the cameras embedded in the mobile eyes of the iCub are fed into an HTC Vive® Head-Mounted Display equipped with an SMI® binocular eye-tracker. The SGCS achieves the effective coupling between the eye-tracked gaze of the pilot and the robot's eye movements. SGCS both ensures a faithful reproduction of the pilot's eye movements – that is perquisite for the readability of the robot's gaze patterns by its interlocutor – and maintains the pilot's oculomotor visual clues – that avoids fatigue and sickness due to sensorimotor conflicts. We here assess the precision of this servo-control by asking several pilots to gaze towards known objects positioned in the remote environment. We demonstrate that we succeed in controlling vergence with similar precision as eyes' azimuth and elevation. This system opens the way for robot-mediated human interactions in the personal space, notably when objects in the shared working space are involved.

1 Introduction

This research is motivated by two main objectives: (a) designing a new generation of immersive control platforms for telepresence robots and (b) teaching multimodal behaviours to social robots by demonstration. In both cases, a human pilot interacts with remote interlocutors via the mediation of a robotic embodiment that should faithfully reproduce the body movements of the pilots while providing rich sensory feedback in coherence with their proprioceptive experience.

During social interactions, people's eyes particularly convey a wealth of information about their direction of attention and their emotional and mental states [2]. Endowing telepresence robots with the ability to mimic the pilot's gaze direction as well as social robots with the possibility to decode, interpret and generate gaze cues is necessary for enabling them to seamlessly interact with humans.

We describe and evaluate here a *beaming* (referring to pioneer work of Slater et al [29]) platform for gaze. This original system goes far beyond our earlier attempts [6], because of the maturing of eye-tracking technology and the use of *foveated rendering* (see below).

The paper is organized as follows. Section 2 overviews the current state of the art on immersive teleoperation of humanoid robots, followed by the description of our technical platform in section 3 and the proposed control method in section 4. Section 5 outlines the experiment we conducted in order to assess the effectiveness of the control method. Results are then presented and discussed in section 6.

Figure 1 Our humanoid robot *Nina* remotely controlled via our teleoperation/beaming platform.

2 State of the art

2.1 Immersive teleoperation

Immersive teleoperation aims at giving an operator the feeling of being present in the remote world and monitoring the robot as its own body, known as *embodiment* [7]. Successful embodiment is achieved when the operator does experience similar proprioception and perception as the robot in the remote environment [12].

Applications for immersive teleoperation of robots include perception-driven tasks (Search and Rescue (SaR) robot [17] and drone navigation [28]), kinesthetic-oriented tasks (immersive telemanipulation [8] and embodied learning [34]), or immersive telepresence [3, 13])

In order to provide an immersive experience, Head-Mounted Displays (HMD) are generally used to provide operators with a stereoscopic first-person view (FPV) of the remote scene. The view is usually provided by a stereo camera attached to the robot's "head". Few systems propose a full control over the human kinematic chain, i.e. independent control of foot, gait, torso, head and eye movements which is so important for the pilot's perception of space and trajectory planning [5] and also for the local human partners' perception of the robot's intentions. Despite the fact that performance and cognitive load are improved by embodied control, few systems [3, 1] actually provide immersive control of the head.

2.2 Immersive teleoperation of head and gaze

The work of Tachi et al [30] and Fernando et al [9] on the *telexistence* cockpit is quite emblematic of immersive

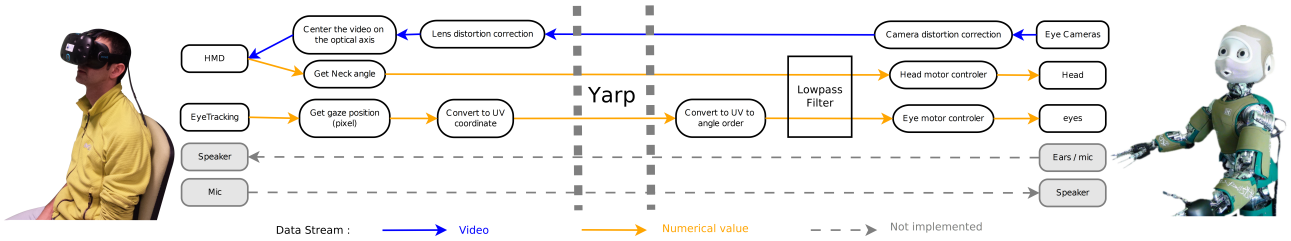

Figure 2 The beaming platform architecture process.

teleoperation of humanoid robots. In their setup, the pilot apprehends the remote environment via several modalities (visual, sound and haptics). In particular, the movements of the controlled HRP robots are steered by his/her head motion while stereo streams from head-related scene cameras are fed back into the HMD displays (see Figure 1).

Recent works on the iCub platform [18] explored immersive teleoperation [15] with full body control using a external Kinect camera [11] or with a data-glove [11, 16]. Alongside these works, Theofilis et al [31] propose a solution to reduce perceived visual latency by dynamically updating an initialized panoramic representation of the distant environment in the VR helmet.

None of the previous works investigated the topic of active and independent teleoperation of eye gaze. The approach proposed by Fritsche et al [11] remains interesting as it expands the angular limitations of the robot neck by using the supplementary degree of freedom of the robotic eyes.

2.3 Gaze-driven teleoperation

Several gaze-driven teleoperation systems have been proposed. As an example, the TeleGaze [14] interface enables the operator to control a robotic platform via eye-gaze tracking supplemented by on-board active vision. The steering of the robot's pan/tilt unit that supports the remote camera is performed by explicit fixations towards specific "action regions" of the interaction screen (look right/left/upwards/downwards). The distinction between free exploration within the current field of view and control of the region is achieved via dwelling time or a simultaneous pedal press. The sensorimotor loop is here totally dislocated. The gaze-driven remote camera control proposed by Zhu et al [35] restores this sensorimotor loop by controlling pan/tilt velocities from the eye-gaze deviations from the center of the screen, while the camera image is streamed directly to the screen. This policy may however lead to conflict between orientation and visual feedbacks.

2.4 Teleoperation of robotic eyes

Immersive teleoperation of the eye movements is even scarcer. While gaze-driven cameras such as the EyeSeeCam [25] have been proposed, no systems combine gaze-driven control and virtual reality (VR) feedback. No wonder, robots with mobile eyes and HMD equipped with embedded binocular eye-trackers are scarce.

The perception of space in current immersive teleoperation devices is affected by several factors, in particular the tunnel-vision problem (caused by limited field of view (FoV) and the well-known accommodation-convergence conflict [32]). This results in the underestimation of depth in the peripersonal space and its overestimation above 2 meters [4]. These limitations are therefore particularly disabling for close face-to-face interactions that involves regions of interest placed in the joint working space.

While mobile eyes allow to increase the scene awareness, we hypothesize that an adequate processing of vergence – at both end: camera capture and HMD display – will lower the accommodation-convergence conflict thanks to an improved consistency between the visual feedback and the oculomotor proprioceptive clues.

3 Our beaming platform

The architecture of our beaming platform is detailed in Figure 2. This platform is hosted by the MICAL technical facility at GIPSA-Lab that consists of two adjacent rooms: (1) a control room that hosts the pilot monitored by motion capture (MoCap) equipments and (2) a Human-Robot Interaction (HRI) room that hosts the robot. The Mocap systems, robot controllers, and processing server units of our local facility are connected by a Gigabit Local Area Network and communicate using the *Yarp* middleware [19] with the UDP or TCP protocols.

3.1 The equipment

Following the work done by Slater et al [27], we beam here the head and eye movements of our iCub2.0 robot endowed with an enhanced articulated talking head [22]. The eyes of this robot have three degrees of freedom (DoF): azimuth/version, elevation/tilt and vergence. Note that the iris texturing and eyelid control [10] is set to ensure a reliable perception of its intended gaze target by human viewers during close face-to-face interaction.

The remote pilot perceives the local scene – in which the robot is located – though a HTC Vive HMD® which has a refresh rate of 90 Hz. The pilot head motion is tracked by the "Lighthouse" base stations with sub-millimeter precision. The pilot eye movements are monitored by an HMD-embedded 250 Hz SMI® binocular eye-tracker. Note that both precision and refresh rates are well above the rates imposed by the robot hardware and/or software: 640x480 images at 30 fps max (closer to 15Hz in practice) and a 100 Hz control loop with a precision of 1° for motors and encoders.

4 SGCS: Stereo Gaze Contingent Steering

Our proposed servo-control (cf. Figure 3) consists of (a) an inverse model that computes angular positions of the robotic eyes given the actual gaze of the pilot captured by the HMD-embedded binocular eye-tracker; (b) a forward model that overlays the "foveal" view delivered by the eye-

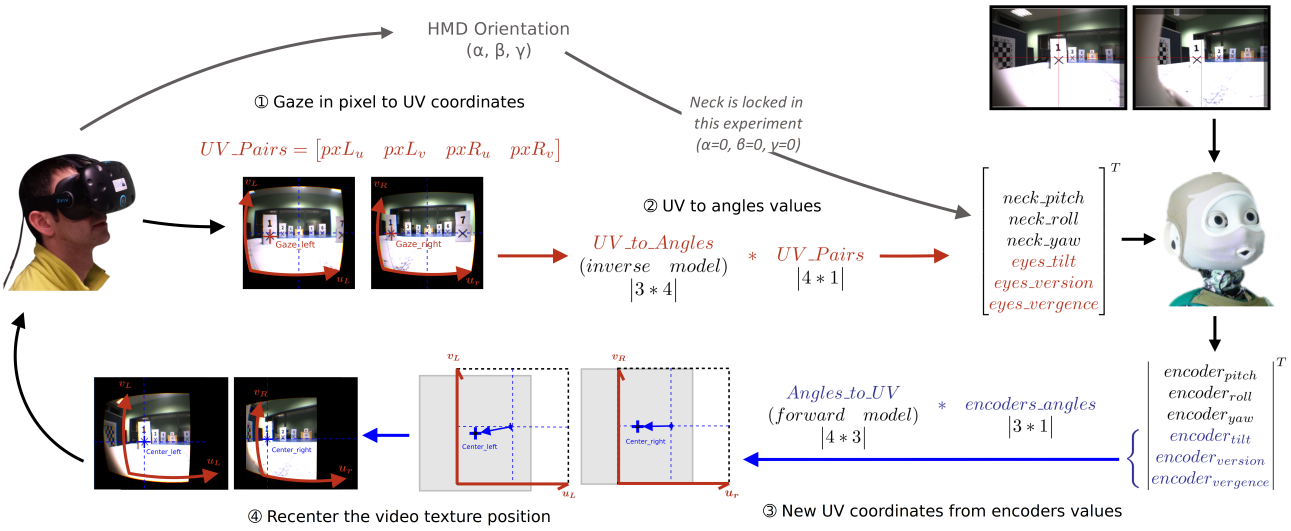

Figure 3 The SGCS control method: (1) The eye-tracking gazing values are converted in UV coordinates. (2) An inverse model computes three angles (azimuth, elevation and vergence) from actual relative gaze coordinates (in UV coordinates). From the values of the motor encoders, (3) a forward model then computes where in the HMD display (4) to overlay the rectified foveal images captured by the cameras embedded into the robot's eyeballs. Note that all sensorimotor transformations but foveal rectification and optical properties of HMD lenses are linear.

embedded cameras of the robot into the field of view of the pilot. Contrary to the *moving to the center* paradigm proposed by Zhu et al [35], the focal images delivered by the cameras are here positioned in the HMD images according to the angular positions of the robotic eyes: it is positioned at the centre when eyes look straight forward (see section 4.1 below), but out of the center otherwise.

4.1 Aligning centers of the fields of view

This sensorimotor system is calibrated in order that the straight ahead gaze of the pilot and the robot align with an object placed at a large distance in the front of the robot. This starting point ensures that the inverse controller generates zero angles for the robot when the pilot is fixating the center of the corresponding left and right video texture displayed in the HMD screen. This is both performed by tuning the orientations of the CCD boards mounted directly on the iCub eye balls and by fine-tuning image translations in the display device. The sensorimotor system has thus a global asymptotically stable equilibrium point at the center of the field of view.

4.2 Linear forward and inverse models

The orientations of the cameras in the eyeballs have thus been set so that angles (*Angles*) set to $[0,0,0]$ correspond to a fixation to a physical target placed at infinity and relative coordinates (UV) in the left and right eye camera images equal to $[0,0]$, i.e. displayed at the center of view of each eye. Note that these camera images will be treated as textures that will be mapped in the HMD screen. The UV range is $[-0.5;0.5]$;

We built a forward and inverse models that map *Angles* to UV and vice-versa. The inverse and forward models (UV_to_Angles and $Angles\_to\_UV$ in figure 3) used here are linear. These transformations are determined thanks to a calibration procedure (only performed off-line once) consisting in moving a pointwise physical target in the field of vision of the robot. Note that UV are determined on rectified images.

For each target position of the calibration step, angles are adjusted (using simple gradient descent) until the target is aligned with each robot's camera optical center. This operation is typically repeated for one hundred targets randomly chosen in the 3D visual cone in front of the robot's head.

Note that the particular contribution of vergence to the regression requires a nonlinear sampling of the visual depth, with more dense samples in the near field ($20-70cm$) than in the far field (i.e. vergence can be considered as null for targets at distance $> 3m$).

Regressions are thus performed between the initial UV of the targets and the final *Angles* at convergence of the alignments.

4.3 Accuracy of the transformations

Whilst we expected this mapping to be highly nonlinear (due to polar transforms and radial rectifications of the images), simple linear regressions surprisingly give quite accurate predictions (see Figures 4 and 5): angular alignment delivered by the *Inverse model* is close to 0.5° while the ROI (region of interest) center delivered by the *Forward model* is reached with median precision of 0.5% in one iteration.

4.4 Mapping robotic gaze shifts with points of interest in the HMD field of view

As in [33], the foveal images captured by the robot's camera are just overlaid into the FoV of the pilot at the relative positions computed by the *Forward model*.

While the cameras have a field of view (FoV) of 63° horizontally and 46° vertically, tilt and version ranges of the eyes increase the HMD viewing area by 100° laterally and 50° vertically. The HMD FoV is thus set at twice the foveal FoV of the cameras: UV coordinates in the HMD span the interval $[-1;+1]$ instead of the original $[-.5;+.5]$ camera range. Note that the range of vertical positions is reduced compared to horizontal ones because of the respective camera models (4/3 ratio for the robot's VGA camera vs. 16/9

for the HMD).

4.5 Using both the inverse and forward models

In order to avoid any inconsistencies between current angular positions – that differ from *Angle* because of the eyes' PID micro-controllers – and stereo video streams delivered by the robot's cameras, encoders of eye angles are read at the moment of video capture. Corresponding positions of the foveal images – that will be displayed into the FoV of the pilot – are then computed thanks to the forward model. We plan to watermark the encoder values in the camera textures in order to secure the consistency between camera positions and contents, but this part of the sensorimotor loop is not so critical: there is no noticeable discrepancy between *What* is displayed and *Where* it is displayed.

4.6 Response time

The measured "average" motion-to-photon (eye movement to corresponding visual feedback) is now around ~200ms. This average response time is due to several lags: (a) the response time of the eyetracker; (b) the response time of the microcontrollers, motors and encoders as well as the friction of the mechanical system of the robotic eyes; (c) the transmission protocol between the motion capture device and the processing modules and also (d) the adaptive filtering of angular trajectories. In fact, we perform a strong low-pass filtering (with a cut-off frequency of 5Hz) of predicted movements during fixations in order to avoid blurring, that would result from delayed responses of microsaccades. On the other end, the cut-off frequency of saccades (i.e. predicted angular movements with a velocity exceeding $300^\circ/s$) is raised at 20Hz.

4.7 Comments

The sensorimotor system thus combines linear models operating around a stable equilibrium point at the center of the field of view. We verified objectively that the sensorymotor loop offers an optimal close-loop convergence: (a) a gaze shift generated with a simulated eye-tracker towards a given point of interest brings it close to the center of the foveal vision of the robot and (b) the cameras embedded in the robot eye balls are effectively directed to the physical target. Because of the linear Inverse and Forward models, the point of interest is positioned after a few iterations (typically 3-5) at the binocular gaze position in the HMD while the robot's eyes point to the physical point in the remote scene. We expect that the imperfections of this extended sensory-motor loop – notably at the limits of the working space – should be easy to compensate by experience.

But this transparent behavior is obtained asymptotically and one might fear that the dynamics of the system (notably motor delays, tracking errors, inertias, etc) may impair this convergence. Moreover, we have no guarantee that binocular accommodation will effectively produce coherent binocular convergence, given known conflicts between accommodation and vergence in existing HMDs [21].

The experiment detailed below in this paper shows that pilots are actually able to implicitly and rightly monitor the vergence of the robot using our system. At our knowledge, it's the first device demonstrating this ability.

5 Evaluation

5.1 Objectives & motivations

The goal of this experiment is to demonstrate that our SGCS paradigm enables the pilot to (1) bring up the robot's eyes to effectively center his/her desired fixation point in the HMD while (2) displaying the camera *foveal* images back at the fixation point.

The first objective enables the distant partners (humans in front of the robot) to decode his/her social gaze. Several studies have effectively shown the strong impact of eye directions and movements on the interpretation of intentions, emotions and attitudes. Foerster et al [10] have further shown that the estimation of the direction of robotic eyes can even outperform that of human eyes when controlled and designed properly.

The second objective aims at providing a robust and precise sensorimotor control while minimizing conflicts between proprioceptive and visual information, notably between accommodation and vergence. It also enforces the beaming experience, with pilot and robot sharing the same gaze states.

5.2 Experiment

In order to characterize our proposed control scheme, we performed a gazing experiment where subjects were asked to gaze at seven static targets, positioned in the robot's field of view as shown in fig. 6. By comparing angular data at fixations to ideal angular configurations collected in 5.2.4, we evaluate if the control method is behaving as expected and confirm that the robot cameras are effectively looking at the same locations as the pilot.

5.2.1 Subjects

Subjects were recruited via an emailing campaign in the lab. Subject were then informed on the tasks, the expected duration of the experiment and the procedure. They were also informed that they can stop the experiment whenever they wanted. Before putting the HMD on, subjects performed two visual tests :

Stereo-blindness Stereo-blindness is defined as the inability to perform stereopsis and perceive stereoscopic depth by combining and comparing images from the two eye [24]. In the "two fingers test", subjects align two fingers in front of their face, with one in the near field, the other in the far field. If while focusing on one finger -and after several trials- the subject still do not perceive multi occurrences of his second finger, he is probably stereoblind. Please note the incidence of stereoblindness in the general population is approximately 2-4% [24]. Stereoblindness was our only rejection criteria as we are not able to calibrate the eye-tracking system in this case and subjects will not verge and accommodate to targets in their near field.

Dominant eye The dominant eye could be defined as the eye that drive the gaze [23]. In order to find out this feature, subjects performed a customized Parson vi-

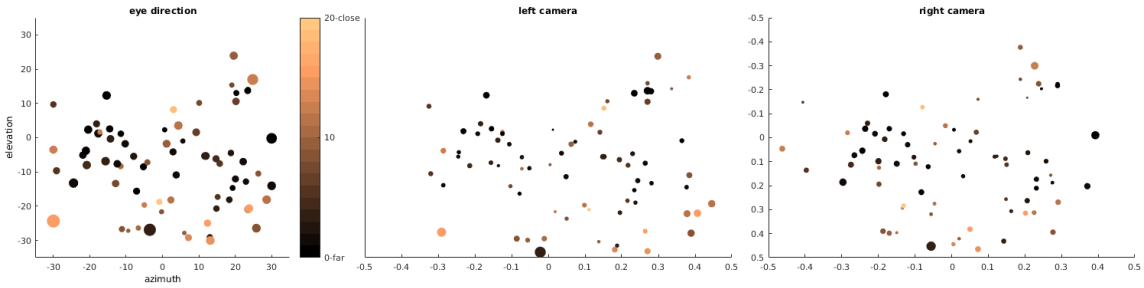

Figure 4 Mapping accuracy of the linear models linking the UV positions of the ROI center of 124 targets – placed randomly in the visual field of view of the robot – in the rectified (l)eft and (r)ight camera images and the relative eye movements required to align this ROI center with the camera axis. Left: the targets are placed in the azimuth/elevation plane. The color features the vergence, i.e. their depth. Diameters of the circles give the average angular error performed by the *inverse model* when fed by the (x,y) positions of the corresponding target in the binocular UV planes displayed at the right. Note that eyelids limit the reachable points of interest at 20° upwards. Right: Reciprocal accuracy of the *forward model*. Diameters of the circles give the average UV error performed by the *forward model* when fed by the ideal angular positions of the targets.

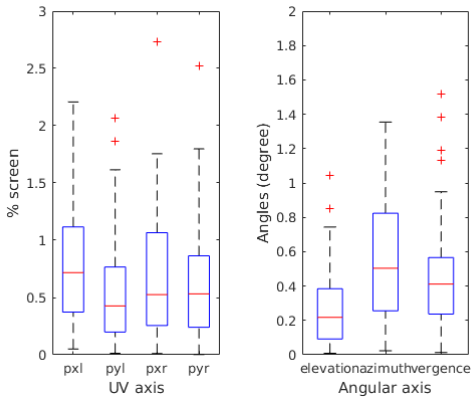

Figure 5 Empirical distributions of the calibration errors displayed in figure 4. The median precision of the angular alignment (right caption) is close to 0.5°. Reciprocally, given the ground truth eye movement, the ROI center (left caption) is reached with median precision of 0.5%. Outliers are mainly due to measurement errors.

sion test [20]. Together with wearing glasses, we expect eye dominance to bias the fixation performance for targets positioned at extreme sides of the field of view.

5.2.2 Calibration of the eyetracker

After putting on and mechanically adjusting the HMD on their face, the subjects performed the standard SMI calibration procedure consisting of a smooth pursuit of a moving calibration target, stopping at a dozen positions spanning the HMD visual field.

5.2.3 Protocol

During the experiment, subjects had to fixate several physical targets - materialized by crosses drawn on cubes – of the remote static scene. All targets are identical: $2.5 * 2.5cm$ tilted black crosses drawn on the white sides of $5 * 5cm$ cubes facing the robot. Each cube is also topped by a number. Subjects were instructed to fixate the center of the cross of the cube whose number was given verbally by the experimenter.

Pilots (subjects) were thus asked to fixate the centers of 7 printed crosses placed in arc in front of the robot (see

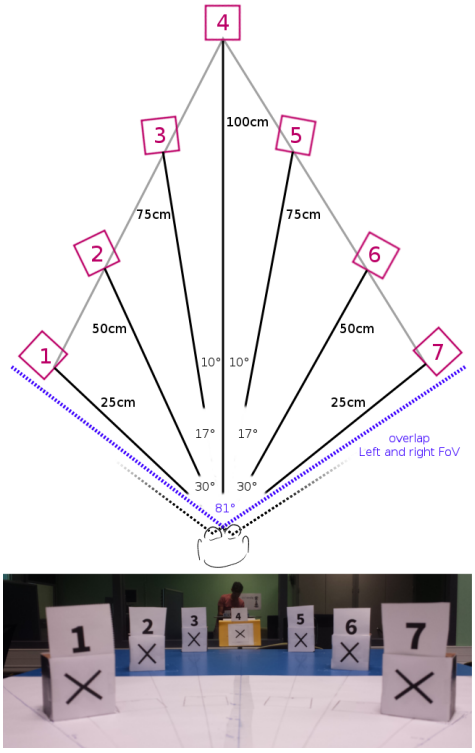

Figure 6 Arrangement of the 7 physical targets used in the validation experiment. We mainly vary the target depths. Pilots are asked to fixate the crosses on the cubes whose target numbers that are instructed verbally.

Figure 6). Each subject browsed the cubes twice in the 4 directions defined in table 1. In order to get asymptotic behaviors, subjects are instructed to hold fixations for at least 2 seconds. At any given moment in time, subjects could blink. They can stop the experiment when wanted. And a break was proposed after the four first task.

Although the actual control paradigm allows the simultaneous control of the orientation of the eyes and the head based on the pilot eyes/head movements (see Figure 3), the robot's head is kept fixed in the present study.

5.2.4 Measurements

The angular commands issued during fixations were sampled at instants of maximum stability. In total, every target was fixated eight times per subject. We thus provide here

the statistics of 104 data points per target over the whole cohort (see figure 7).

Ground-truth measurements were performed (in magenta in figure 7): the angular position of the cameras are set in order to position each target precisely at the center of each camera image. Visual crosses are overlaid on the video feedbacks of the operator to verify that their centers match the targets.

6 Results

13 pilots (3 women, 10 men) with normal (10) or corrected vision (3) and no prior experience in VR participated in the experiment. Fixations and control measurements are displayed in the Figure 7. Angular values of the robotic eyes at the fixations (yellow distributions) are centered close to the control measurements (magenta positions): all pilots actually fixate the targets with angular commands which are very close to the ideal ones.

6.1 Objective results

Statistical tests show that the distributions of angular errors do solely depend on the target positions but not on the serial ordering of the screening. Posthoc Tukey HSD tests show that distributions of elevation and azimuth errors only differ for t7 ($p < .001$) while distributions of vergence only differ for t1 ($p < .001$).

Eye fixations of the subjects resulting from the SGCS are effectively directed towards the 7 targets whatever their serial ordering (see Figure. 8). We studied the dispersion of eye movements in the vicinity of fixations. Gaussian mixture models (GMM) were fitted to raw left and right gaze positions monitored by the eyetracker vs. target positions of the visual feedback positions computed by the forward model from the controlled angles of the robot's eyes.

Dispersion ellipsis have comparable shapes and size and attest that, when the pilot is looking at a target, the eyes are exactly gazing at the same location: the cameras' vision axis and human eye gaze direction are aligned.

6.2 User feedbacks

At the end of experiments, an informal oral survey was conducted in order to subjectively evaluate the control system. Most of the subjects indicated that they had a good sense of presence in the distant world, appreciated response time of saccades and did not report being hindered by the photon-to-movement lag for this experiment. They however reported difficulty to focus on extreme targets, as they saw them as blurred and could not move their head. They also indicated eyestrain at the end of the session (certainly due to the HMD back-lighting). Two subjects reported a slight cybersickness.

Order	Target verbal instructions
left-to-right	$1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 5 \rightarrow 6 \rightarrow 7$
front-to-back	$1 \rightarrow 7 \rightarrow 2 \rightarrow 6 \rightarrow 3 \rightarrow 5 \rightarrow 4$
right-to-left	$7 \rightarrow 6 \rightarrow 5 \rightarrow 4 \rightarrow 3 \rightarrow 2 \rightarrow 1$
back-to-front	$4 \rightarrow 3 \rightarrow 5 \rightarrow 2 \rightarrow 6 \rightarrow 1 \rightarrow 7$

Table 1 Each subject (pilot) browsed the 7 cubes in four different orders (left column). This is performed twice per pilot. The instructions are given verbally.

6.3 Discussion

We show that the servo-control is effective in monitoring the 3 DOFs of the robotic eyes. We are notably very enthusiastic by the performance of the steering of the vergence. At our knowledge, it is the first time that such a performance is obtained.

Several methodological and technological bolts have still to be resolved to conduct pleasant and fluid robot-mediated interactions:

Head control. The evaluated configuration of the SGCS was "head-locked". Even if our gaze control method already combines head and eye teleoperation, a complementary evaluation should be conducted to evaluate the true potential of SGCS in a head-free exploration and perception task.

Dynamic behavior. If the quasi-static precision of fixations is quite satisfactory, saccades and smooth pursuits are still impaired by the sensorimotor delay induced by the hardware/software components (now close to 60 ms in absence of low-pass filtering). While the eye-tracker and micro-controllers run at 90 Hz and 100 Hz respectively, VGA image capture is typically limited to 20-30 images per seconds in the special iCub2.0 version with Atom CPU. Predictive coding and machine vision might possibly be combined to lock intended targets of fixations and smooth pursuits.

Limits of the FoV. The loss of precision at t1 and t7 may be explained by image deformations due to lenses of the cameras and the HMD. Another issue is that internals of the robotic eyes might be seen by one of the two cameras when turning fully to the right or fully to the left, which breaks the 3D reconstruction done by the brain as it's not coherent across the two eyes. The use of such extreme positions is reduced when the teleoperation of the head orientation is enabled, allowing head and eyes to move in synergy. The kinematic chain of audiovisual attention recruits the whole body – from the toes to the eyes – and we expect a full body beaming to even better solve these issues.

Peripheral vision. The foveal vision should be complemented by peripheral vision so that exogenous stimuli arising away from the limited FoV of the cameras may attract attention and trigger saccades with proper amplitudes. This could be achieved by blending the output of a scene camera with the foveal textures (see the sharp gaze spot of the Eyeseecam project [26]).

7 Conclusions

We describe here a beaming platform that enables pilots to seamlessly monitor the head and eye gaze movements of an embodied robotic platform, while maintaining the coherence between oculomotor cues and visuo-kinesthetic feedback. This coherence should benefit to the pilot's sense of embodiment and presence in the remote environment. The control loop is reliable and faithful. At your knowledge, it is the first system to successfully control eye vergence by immersive teleoperation.

This performance opens the way to numerous applications, including teaching social behaviors by demonstration.

There is still a large room for improvements. For exam-

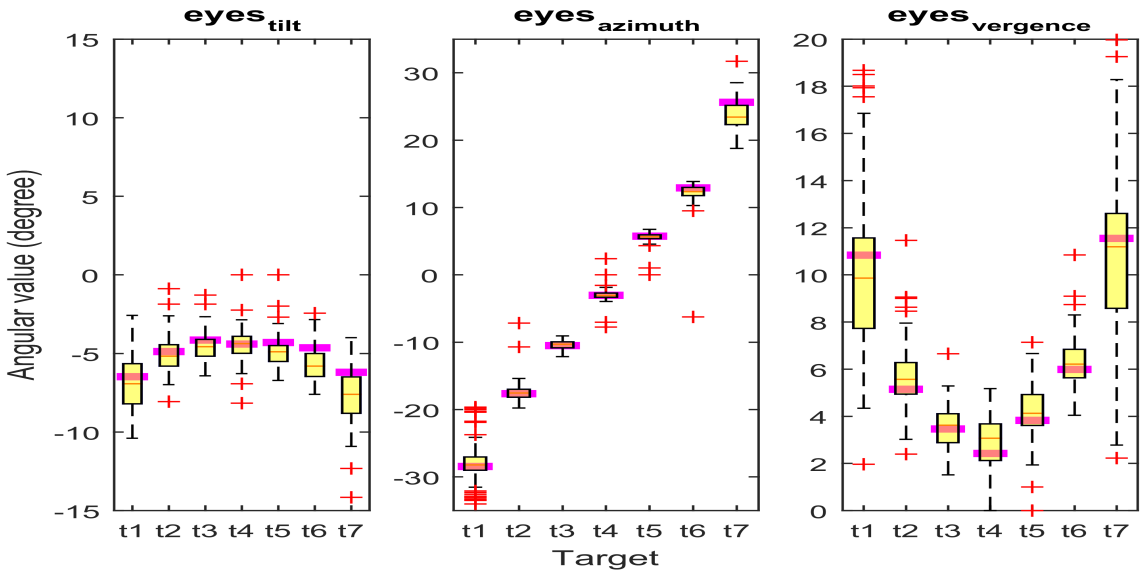

Figure 7 Realized (yellow boxplots) vs. control (magenta) ocular angles for the different cubes. From left to right: elevation, azimuth and vergence. Each boxplot figures 104 data points (13 pilots x 8 fixations per target).

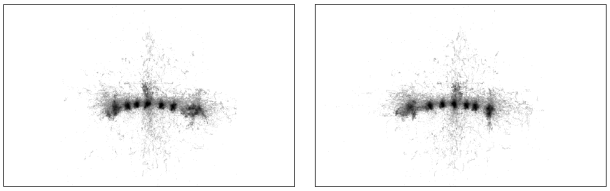

Figure 8 Raw distribution eye-tracking data of the left vs. right eyes of all subjects (HMD *UV*). The clustering of gaze positions around 7 clusters – that indeed correspond to the target cubes – is not trivial since it is the result of a sensorimotor control that projects back camera images in mobile positions computed as a function of *Angles*, that may differ from current fixation of the current subject's gaze due to delays and modeling errors of the inverse and forward kinematics. Note that the active HMD FoV is larger than the fixed camera FoV because of the mobility of the cameras.

ple, the foveal and parafoveal information is limited with the present mobile cameras and could be complemented by an additional head-mounted camera with a wide FoV. The dynamics (cut-off frequency, data synchronization, delays, etc) of the control should also be enhanced before the system may be used by untrained pilots for high-demanding robot-mediated interactive tasks.

Based on these results, we now have a coherent oculomotor and vestibular representation of the remote world in the HMD, which enhanced the feeling of being embodied [12]. Future evaluation will be pursued on the pilot in order to quantify body ownership and self-location, notably by exploring the impact of SGCS on depth perception.

Acknowledgments

This work is supported by ROBOTEX (ANR-10-EQPX-44-01), SOMBRERO (ANR-14-CE27-0014), PERSYVAL (ANR-11-LABX-0025) and by the TENSIVE project.

Literature

[1] Evan Ackerman. 2015. Oculus Rift-Based system brings true immersion to telepresence robots. *IEEE Spectrum* (2015).

[2] Reginald B Adams Jr and Robert E Kleck. 2003. Perceived gaze direction and the processing of facial displays of emotion. *Psychological science* 14, 6 (2003), 644–647.

[3] Luis Almeida, Paulo Menezes, and Jorge Dias. 2017. Improving robot teleoperation experience via immersive interfaces. In *IEEE Experiment@ Int. Conf.* 87–92.

[4] C Armbrüster, M Wolter, T Kuhlen, W Spijkers, and B Fimm. 2008. Depth perception in virtual reality: distance estimations in peri- and extrapersonal space. *Cyberpsychology & behavior : the impact of the Internet, multimedia and virtual reality on behavior and society* 11, 1 (2008), 9–15.

[5] Colas N Authié, Pauline M Hilt, Steve N'Guyen, Alain Berthoz, and Daniel Bennequin. 2015. Differences in gaze anticipation for locomotion with and without vision. *Frontiers in human neuroscience* 9 (2015).

[6] Gérard Bailly, Frédéric Elisei, and Miquel Sauze. 2015. Beaming the gaze of a humanoid robot. In *ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI) Extended Abstracts*. 47–48.

[7] F. Biocca. 1997. The cyborg's dilemma: embodiment in virtual environments. *IEEE Int. Conf. on Cognitive Technology Humanizing the Information Age* (1997), 12–26.

[8] Junshen Chen, Marc Glover, Chenguang Yang, Chunxu Li, Zhijun Li, and Angelo Cangelosi. 2017. Development of an immersive interface for robot teleoperation. In *Lecture Notes in Computer Science (LNAI)*. Vol. 10454. Springer, 1–15.

[9] Charith Lasantha Fernando, Masahiro Furukawa, Tadatoshi Kurogi, Sho Kamuro, Kouta Minamizawa, Susumu Tachi, and others. 2012. Design of TELESAR V for transferring bodily consciousness in telexistence. In *IEEE Int. Conf. on Intelligent Robots and*

Systems (IROS). 5112–5118.

[10] François Foerster, Gérard Bailly, and Frédéric Elisei. 2015. Impact of iris size and eyelids coupling on the estimation of the gaze direction of a robotic talking head by human viewers. In *IEEE Int. Conf. on Humanoid Robots (Humanoids)*. 148–153.

[11] Lars Fritsche, Felix Unverzag, Jan Peters, and Roberto Calandra. 2015. First-person tele-operation of a humanoid robot. In *IEEE-RAS Int. Conf. on Humanoid Robots (Humanoids)*. 997–1002.

[12] Konstantina Kilteni, Raphaela Groten, and Mel Slater. 2012. The sense of embodiment in virtual reality. *Presence: Teleoperators and Virtual Environments* 21, 4 (2012), 373–387.

[13] S. Kratz and F. Rabelo Ferriera. 2016. Immersed remotely: Evaluating the use of Head Mounted Devices for remote collaboration in robotic telepresence. In *IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. 638–645.

[14] Hemin Omer Latif, Nasser Sherkat, and Ahmad Lotfi. 2009. Teleoperation through eye gaze (TeleGaze): a multimodal approach. In *IEEE Int. Conf. on Robotics and Biomimetics (ROBIO)*. 711–716.

[15] Uriel Martinez-Hernandez, Luke W Boorman, and Tony J Prescott. 2015. Telepresence: immersion with the iCub humanoid robot and the Oculus Rift. *LNCS* 9222 (2015), 461–464.

[16] Uriel Martinez-Hernandez, Luke W Boorman, and Tony J Prescott. 2017. Multisensory Wearable Interface for Immersion and Telepresence in Robotics. *IEEE Sensors Journal* 17, 8 (2017), 2534–2541.

[17] Henrique Martins, Ian Oakley, and Rodrigo Ventura. 2015. Design and evaluation of a head-mounted display for immersive 3D teleoperation of field robots. *Robotica* 33, 10 (dec 2015), 2166–2185.

[18] Giorgio Metta, Giulio Sandini, David Vernon, Lorenzo Natale, and Francesco Nori. 2008. The iCub humanoid robot: an open platform for research in embodied cognition. *Workshop on Performance Metrics for Intelligent Systems* (2008), 50–56.

[19] Metta, Giorgio, Paul Fitzpatrick, and Lorenzo Natale. 2013. YARP – Yet Another Robot Platform, version 2.3.20. *International Journal of Advanced Robotic Systems* 3, 1 (2013), 43–48.

[20] Walter R. Miles. 1930. Ocular Dominance in Human Adults. *The Journal of General Psychology* 3, 3 (jul 1930), 412–430.

[21] Jincheol Park, Sanghoon Lee, and Alan Conrad Bovik. 2014. 3D visual discomfort prediction: vergence, foveation, and the physiological optics of accommodation. *IEEE Journal of Selected Topics in Signal Processing* 8, 3 (2014), 415–427.

[22] Alberto Parmiggiani, Marco Randazzo, Marco Maggiali, Giorgio Metta, Frederic Elisei, and Gerard Bailly. 2015. Design and Validation of a Talking Face for the iCub. *International Journal of Humanoid Robotics* 12, 03 (2015), 1550026.

[23] Clare Porac and Stanley Coren. 1976. The dominant eye. *Psychological Bulletin* 83, 5 (1976), 880–897.

[24] Whitman Richards. 1970. Stereopsis and stereoblindness. *Experimental Brain Research* 10, 4 (1970), 380–388.

[25] Erich Schneider, Thomas Dera, K Bard, Stanislavs Bardins, Guido Boening, and T Brand. 2005. Eye movement driven head-mounted camera: it looks where the eyes look. In *IEEE Int. Conf. on Systems, Man and Cybernetics*, Vol. 3. IEEE, 2437–2442.

[26] Erich Schneider, Thomas Villgrattner, Johannes Vockeroth, Klaus Bartl, Stefan Kohlbecher, Stanislavs Bardins, Heinz Ulbrich, and Thomas Brandt. 2009. Eyeseecam: An eye movement–driven head camera for the examination of natural visual exploration. *Annals of the New York Academy of Sciences* 1164, 1 (2009), 461–467.

[27] Mel Slater, Anthony Steed, and Martin Usoh. 2013. *Being There Together: Experiments on Presence in Virtual Environments (1990s)*. Vol. 6. Springer-Verlag John Wiley and Sons. 1–218 pages.

[28] Nikolai Smolyanskiy and Mar Gonzalez-Franco. 2017. stereoscopic First Person View system for Drone navigation. *Frontiers in Robotics and AI* 4 (2017), 1–10.

[29] Anthony Steed, William Steptoe, Wole Oyekoya, Fabrizio Pece, Tim Weyrich, Jan Kautz, Doron Friedman, Angelika Peer, Massimiliano Solazzi, Franco Tecchia, and others. 2012. Beaming: an asymmetric telepresence system. *IEEE computer graphics and applications* 32, 6 (2012), 10–17.

[30] Susumu Tachi, Kiyoshi Komoriya, Kazuya Sawada, Takashi Nishiyama, Toshiyuki Itoko, Masami Kobayashi, and Kozo Inoue. 2003. Telexistence cockpit for humanoid robot control. *Adv. Robotics* 17, 3 (2003), 199–217.

[31] Konstantinos Theofilis, Jason Orlosky, Yukie Nagai, and Kiyoshi Kiyokawa. 2016. Panoramic view reconstruction for stereoscopic teleoperation of a humanoid robot. In *IEEE-RAS Int. Conf. on Humanoid Robots*. 242–248.

[32] Margarita Vinnikov and Robert S. Allison. 2014. Gaze-Contingent Depth of Field in Realistic Scenes: The User Experience. In *ACM Symposium on Eye Tracking Res. and Appl. (ETRA)*. 119–126.

[33] Johannes Vockeroth, Thomas Dera, Guido Boening, Klaus Bartl, Stanislavs Bardins, and Erich Schneider. 2007. The combination of a mobile gaze-driven and a head-mounted camera in a hybrid perspective setup. In *IEEE Int. Conf. on Systems, Man and Cybernetics*. 2576–2581.

[34] Tianhao Zhang, Zoe McCarthy, Owen Jow, Dennis Lee, Ken Goldberg, and Pieter Abbeel. 2018. Deep Imitation Learning for Complex Manipulation Tasks from Virtual Reality Teleoperation. In *IEEE Int. Conf. on Robotics and Automation (ICRA)*.

[35] Dingyun Zhu, Tom Gedeon, and Ken Taylor. 2011. "Moving to the centre": A gaze-driven remote camera control for teleoperation. *Interacting with Computers* 23, 1 (jan 2011), 85–95.